

AG

METHOD AND DEVICE FOR TRANSLATION

Patent Number: JP63037471

Publication date: 1988-02-18

Inventor(s): OKAJIMA ATSUSHI

Applicant(s): HITACHI LTD

Requested Patent: JP63037471

Application Number: JP19860179919 19860801

Priority Number(s):

IPC Classification: G06F15/38

EC Classification:

Equivalents:

Abstract

PURPOSE: To facilitate a user's proofreading and to improve the efficiency of proofreading work by evaluating the user's processing while a device is translating, separating a correctly translated part and an unreliable part and displaying them.

CONSTITUTION: A first language and proofreading information are inputted to a CPU 3 from a keyboard 1 in a translating machine, and a translated sentence and evaluation information are displayed on a display device 2. A main memory 4 connected to the CPU 3 stores a variety of information, and a dictionary file 5 stores 1st and 2nd languages by comparing them. The CPU 3 controls the memory and the file. A processing program 41 in the memory 4 executes algorithm for translation, and rules for translation are recorded in a translation rule table 42. Originals and translated sentences are stored in an I/O text area 43, thresholds to evaluate translated sentences in a threshold table 44, and the evaluated results in a partial translation result area 45. The sentences are divided into the correctly translated part and the unreliable part, which are displayed on the device 2.

Data supplied from the esp@cenet database - I2

⑫ 公開特許公報 (A)

昭63-37471

⑬ Int.C1.
G 06 F 15/38識別記号
厅内整理番号
7313-5B

⑭ 公開 昭和63年(1988)2月18日

審査請求 未請求 発明の数 2 (全5頁)

⑮ 発明の名称 翻訳方法及び装置

⑯ 特願 昭61-179919

⑰ 出願 昭61(1986)8月1日

⑮ 発明者 岡島 悅 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

⑯ 出願人 株式会社日立製作所 東京都千代田区神田駿河台4丁目6番地

⑰ 代理人 弁理士 小川 勝男 外1名

明細書

1. 発明の名称

翻訳方法及び装置

2. 特許請求の範囲

1. 第1言語から第2言語への翻訳処理において、翻訳時の自己の処理を評価し、一定以上正しいと考えられる部分と、他とを分離してその訳語を表示し、各々これを校正することを特徴とする翻訳方法。

2. 翻訳の為に必要な辞書ファイルと、これを利用して翻訳する手段と、翻訳処理を評価する手段と、翻訳結果と評価結果を格納する記憶手段と、上記評価結果に従い、第1言語の入力文を分割し、各部分に対応する訳文を表示する表示手段と、該表示結果を校正する校正手段とを有することを特徴とする翻訳装置。

3. 発明の詳細な説明

〔達成上の利用分野〕

本発明は、自然言語で表現された文を翻訳するための装置及び方式に関するものである。

対象となる言語は任意であるが、便宜上、主に英語を日本語に翻訳する場合について述べる。

〔従来の技術〕

従来、自然言語で表現された文を機械で翻訳する方式を、翻訳や訳文の多義に対する処理方式の面から見てみると、以下のように大別できる。

(1) 第1言語の解析の成功・失敗にかかわらず第2言語の訳文を作る。

(2) 解析が失敗したことを表示する。

(この場合、訳文を出力する場合としない場合がある。)

(3) さらに、失敗はしないが、あやしい場合を含め、どの程度確信度を持つて翻訳したかとか、どのようなあやしさかを表示する。

上記の(2)で訳文を出力しない場合を除いては、何らかの訳文を出力することになるが、従来の方式では、以下のようないくつかの問題がある。

上記(1)の方式では、ユーザにとつては、正しい訳と誤った訳が、どのように出力されるかわからぬため、一字一句に注意して訳文や原文を見な

ければならず、時には、正しい訳までも戻つて読み直すということも起る。

上記(2)の方法では、解析失敗の表示があるものについては注意して見ることにより校正能率が上がるが、解析が失敗又はあやしいと表示されても、どの程度のあやしさかとか、どのような種類のあやしさかが不明のため、原因を予測し正しく校正するのに手間がかかる。

上記(3)の方法では、(2)の欠点を解消する。その一例は、本発明の同一出願人による特開昭60-86331があり、それによれば、単位は、あらかじめ原文から決つているような文といつた翻訳単位でなく、句・節といつた構文解析して初めてわかるものの単位で表示することもできるため、校正がしやすい。しかし、その場合でも、正しい訳とあやしい訳とが混つていることは避けられず、原文との対応等に手間取る。この欠点を解消するためには、訳語内的一部のみを出力するという方法が考えられるが、その場合も、動詞だけ出力するとか名詞だけ出力する等の方式とすると、英語

上記目的は、翻訳処理の過程で、自己の処理に対する評価をする機構と、この評価機構により、正しく翻訳されたと思われる部分とあやしいと思われる部分とを分離して表示し、これをユーザが各自校正し、まとめるにより全部を作成するような校正装置により達成される。

【作用】

自己評価部は、翻訳処理の各過程が、どれくらいの確信度で行なわれたかと、どのような種類のあやしさがあるかを把握する。この結果に従い、部分毎に確信度に従いユーザに翻訳結果が表示されるので、ユーザは、確信度の高い部分を核にして、他の部分を使つて全文を組立てることができるので、正しい訳文と誤訳が混在して表示される場合に比べ、能率良く校正ができる。

【実施例】

以下、本発明の一実施例を図面に基づいて説明する。

第2図は、本発明の一実施例による翻訳装置のブロック図である。

のような多品詞語の場合は、名詞が動詞に誤解析されたような場合にはかえつて、主要語を落してしまうことになりかねない欠点がある。

【発明が解決しようとする問題点】

上記従来技術は、機械翻訳の校正処理において、正しく翻訳された部分と翻訳処理があやしい部分が混ると校正者にとつては、それら相互の分離や誤りの識別が難しいことに対する配慮がされておらず、それらへの対処に時間がかかるつたり、正しい部分をも戻つて読み直さなければならないという精神的負担の問題があつた。

本発明の目的は、上記の問題に対し、翻訳処理の過程で判明する句・節といつた部分毎に、どのような種類のあやしさが、どの程度あるかを自己評価し、さらに、人間にとつては、長文全体の誤りや不具合を見つけるのは難しいのに、短い部分では、かなりを誤りがあつても見つけ易いという性質を利用して能率よく校正作業をすることにある。

【問題点を解決するための手段】

第2図において、1は第1言語の入力や校正情報等を入力するためのキーボード、2は翻訳文や評価情報等を表示するための表示装置、3はCPU、4は各種情報を格納しているメインメモリ、5は、第1言語と第2言語とを対比して格納している辞書ファイルである。

第3図は、第2図に示したメインメモリ4の割付け配図である。

メインメモリ4は、翻訳のためのアルゴリズムを実行する処理プログラム41、翻訳のための規則を記憶している翻訳用規則テーブル42、原文および訳文を格納していく入出力テキストエリア43、翻訳対象文の多義性に従い確信度を計算した際、これをどのように評価するかを決めるしきい値表44、各部分の翻訳結果とその評価結果を格納する部分翻訳結果エリア45、訳文を校正するための部分訳文校正用エリア46、ワークレジスター47を有する。

次に本実施例の処理動作について説明する。本実施例は、主に英語を日本語に翻訳する英日翻訳

について示したが、他の言語間の翻訳についても本発明が適用可能ることは言うまでもない。

ユーザの指示により翻訳処理が開始されると、CPU 3は、処理プログラム 4 1を起動する。

第1図は、本発明の一実施例におけるプログラム 4 1の処理動作を示すフローチャートである。

処理は、まず第1言語で書かれた文章を辞書ファイル 5を参照しながら構文解析する。この際、解析中に各部分毎に解析の確信度をつける(101)。この方式は一例としては本発明の同一出願人による特願昭60-86331等によることが出来る。

次に各部分の訳文を生成する(102)。上記101, 102の翻訳の方式は、特に規定しないが、同一出願人に係る発明特願昭56-138586等によることが出来る。

上記により、入出力テキストエリア 4 3には、原文と訳文が、又部分翻訳結果エリア 4 5には部分に分けられた原文(第1言語)とその訳文(第2言語)および評価結果が格納される。

次に、原文と訳文をある確信度(これをしきい

値表4 4に格納し、これと比較する)以上の部分とその他の部分とに分けて(混ぜないで)表示装置 2に表示する(103)。

その一表示例を第4図に、又、部分翻訳結果エリア 4 5の一例を第5図に示す。第4図は、訳文の内、一番確信度の高い訳文を、訳の主文とし、その他を校正の為の部品として出力した形となっている。

表示が終ると、ユーザは、訳文として必要な部分を、各部分毎に校正し、これを合成するような入力をを行い、処理プログラム 4 1は、このユーザ入力により校正処理を行なう(104)。

第4図の例をとれば、第2、第3の部分の語尾を修正し、訳の多義を選択し、その後、第1の部分に、これらを合成させる。合成の指示は、第6図の例の32, 33のような部分を示すコードを主となる部分に追加する方法でも、32の部分をカーソルで探し機能キーで追加を指示するといった方式でも良い。上記の指示により、第7図に示すような正しい訳が得られる。

〔発明の効果〕

以上説明したように、本発明の翻訳機によれば、正しく翻訳されているであろう部分とあやしい部分とが分離表示されるため、等に長文に対して、まず正しいと思われる部分の訳文を読み、文全体の正しい概要を捉んだ上で、その他よりあやしい部分の訳を部品として利用して文を作り上げることにより、正しい部分とあやしい部分が混存して表示される場合に起る、思考の混乱や、さらには正しい部分をも表つて原文を読み直すといった無駄をなくすことが可能となり、能率の良い翻訳文の校正を行なうことができる。

さらに、文の主となる動詞(主動詞)の解析に重点を置いて翻訳規則を作成し、その部分を中心又は先に解析するようにし(主動詞解析)、この部分を中心(例えば、確信度がこれより高い部分があつても、主文を中心に)校正を行なうことも可能で、その場合、主文以外の部分を最終テキストに残さなければ、主文を中心に原文のアブストラクトを作るシステムを実現することができる。

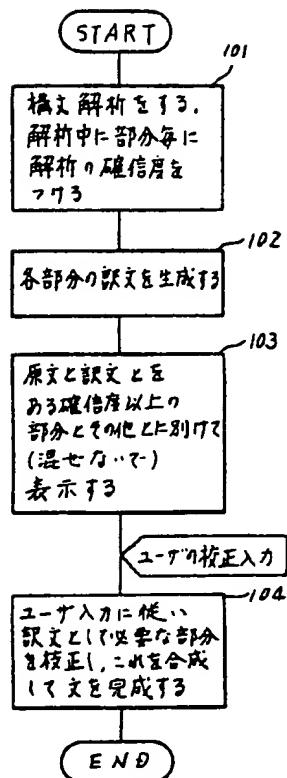
4. 図面の簡単な説明

第1図は本発明方式の一実施例を示すフローチャート、第2図は本発明の一実施例による翻訳装置のブロック図、第3図は第2図におけるメインメモリの割付け配置図、第4図は本発明における校正の為の訳文および部分訳の表示例の一例を示す図、第5図は本発明による部分翻訳結果エリアの一例を示す図、第6図は本発明による部分翻訳結果を合成する指示方式の一例を示す図、第7図は第6図の指示により完成された訳文の一表示例を示す図である。

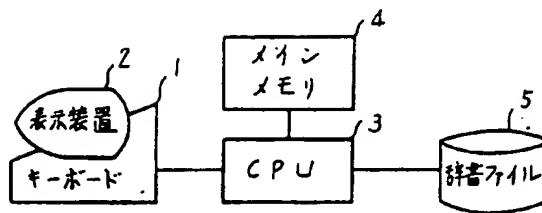
1…キーボード、2…表示装置、3…CPU、4…メインメモリ、5…辞書ファイル。

代理人弁理士 小川勝男

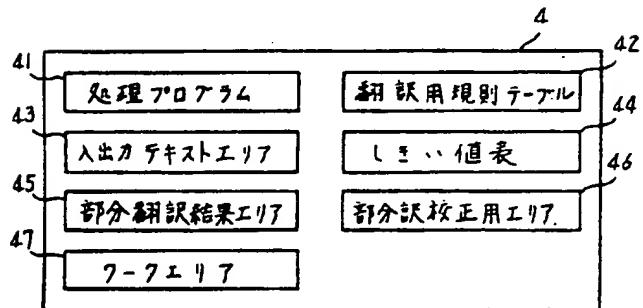
第 1 図



第 2 図

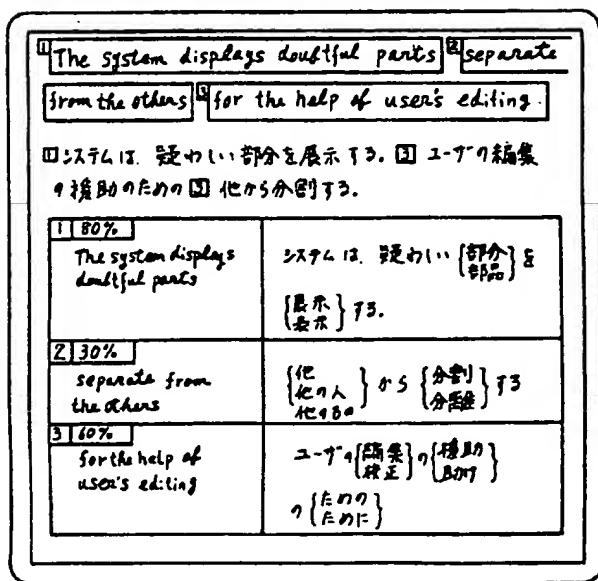


第 3 図



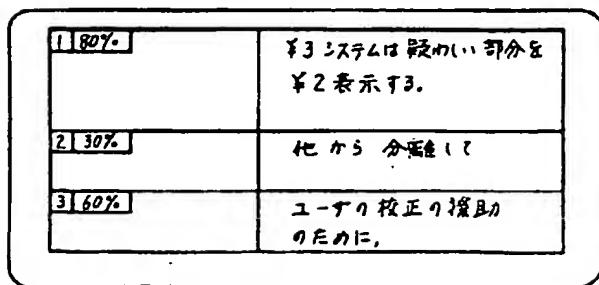
第 5 図

第 4 図

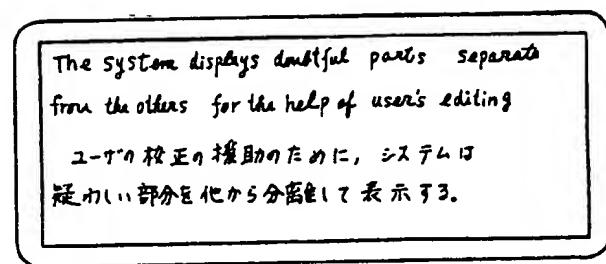


部分番号	確信度	原文	訳文
1	80	The system displays doubtful parts	システムは、疑わしい部分を表示する。
2	30	separate from the others	他から分離する。
3	60	for the help of user's editing	ユーザの校正の手助け

第 6 図



第 7 図



1/31

File with #261

ABSTRACT:

A software program utility that isolates translatable English text from machine code, and merges the text with translated text into one file for review purposes.

DESCRIPTION OF THE PROBLEM SOLVED:

Our team translates RSS products into other languages. After the return of the data from world-wide translation centers, there is an independent local review process to verify the accuracy of the translation between English and language text such as Spanish, French, or German.

The problem was a need to find a more efficient method for reviewers to review our products. Difficulties the reviewers faced included:

1. The inability to isolate reviewable text from source code, and
2. The cumbersome process of simultaneously managing both the English and language files and to keep them in sync for the review.

for non
NL-enabled
code
- since NL code
code creates many
txt file + handles
accordly.
- Batch files have
3rd. part in source.

HOW THE INVENTION SOLVES THE PROBLEM:

My invention isolates English text from source code, merges it with a copy of translated text, and writes the combination into one file.

Instead of having two side-by-side workstation screens or two hardcopy documents interspersed with source code, the reviewer has one soft or hardcopy document, with pairs of text strings. Each entry pair contains only English text and translated text, and no source code.

how find English? - look up tables - - correction - 261. - test & press -

top-bottom
identifies file was
reference # for
editing
correction instead of
inventi.

OTHER METHODS OF SOLVING THE PROBLEM:

I know of no other method to solve the problem other than the former method. The former method involved presenting the English and language text as:

- a) two side-by-side workstations with the data on the two screens, or
- b) one screen with a "toggling" facility, or
- c) two hardcopy documents

For hardcopy documents, we have attempted to isolate the reviewable text from the source code using highlighting pens. There was no practical method to isolate the text from the code when softcopy was presented.

- Xlations do not affect code of source which is non-essential language.
- Verify Xlations of comments lines.
- Source code presence complicates viewer's review.
- Compare / stripper.